

Comparing and generating Latin Hypercube designs in Kriging models

Original

Comparing and generating Latin Hypercube designs in Kriging models / Pistone, Giovanni; Vicario, Grazia. - In: ASTA ADVANCES IN STATISTICAL ANALYSIS. - ISSN 1863-8171. - 94:(2010), pp. 353-366. [10.1007/s10182-010-0142-1]

Availability:

This version is available at: 11583/2394654 since:

Publisher:

Springer

Published

DOI:10.1007/s10182-010-0142-1

Terms of use:

openAccess

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

Comparing and generating Latin Hypercube designs in Kriging models

Giovanni Pistone · Grazia Vicario

Received: date / Accepted: date

Abstract In Computer Experiments (CE), a careful selection of the design points is mandatory to predict the system response at untried points, based on the values observed at tried points. In physical experiments, the protocol is based on Design of Experiments, a methodology whose basic principles are questioned in CE. When the responses of a CE are modeled as jointly Gaussian random variables with covariance depending on the distance between points, the use of the so called space-filling designs (random designs, stratified designs and Latin Hypercube designs) is a common choice, because it is expected that the nearer untried point is to the design points, the better is the prediction. In this paper we focus on the class of Latin Hypercube (LH) designs. The behavior of various LH designs is examined according the Gaussian assumption, in order to minimize the total prediction error at the points of a regular lattice. In such a special case, the problem is reduced to an algebraic statistical model which is solved by using both symbolic algebraic software and statistical software. We provide closed-form computation of the variance of the Gaussian linear predictor as a function of the design, in order to make a comparison between LH designs. In principle, the method applies to any number of factors and any number of levels, and also to classes of designs other than LHs. In our current implementation, the applicability is limited by the high computational complexity of the algorithms involved.

Keywords Computer Experiments · Latin Hypercube · Mean Square Prediction Error · Gaussian linear prediction

Mathematics Subject Classification (2000) 62K99

The first Author acknowledges the support of DIMAT Politecnico di Torino and of SAMSI, NC, he visited Jan-Mar 2009, where the computational part was developed and some issues were discussed in a working group of the 2008-09 SAMSI Program on Algebraic Methods in Systems Biology and Statistics. The second Author acknowledges the support of DISPEA Politecnico di Torino

DIMAT Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino Italy E-mail: giovanni.pistone@gmail.com · E-mail: grazia.vicario@polito.it

1 Introduction

The official start of Computer Experiments (CE) the paper by McKay et al (1979), while the contribution by Sacks et al (1989b) marked a new step by introducing model-based methods. A compelling reason for using CE (in a single or a combined approach with the physical experiments) comes from the fact that physical experimentation may be, in a number of circumstances, expensive or even unapproachable. On the contrary, the use of numerical experiments in product/process development phase is relatively inexpensive and, because of that, has become straightforward. The general availability of comprehensive computing facilities and the recent progresses in software development make numerical simulation of complex systems an attractive alternative option to the execution of the expensive and time consuming physical experiments. Standard modern references are Sasena (2002), Santner et al (2003), Fang et al (2006).

In this context, a careful selection of the design points or training points is mandatory to predict how the unobserved responses depend on the observed ones. In physical experimentation, the researcher is asked to comply to a well set protocol in order to achieve correct inferences. Such a protocol is the Design of Experiments (DoE) methodology which is an helpful tool in carrying on the mentioned objectives. The design of a CE, when it is used as a surrogate of the physical one, differs in several aspects from designing a physical experiment and the applicability of basic principles of DoE is questioned in CE. The selection of an experimental design in CE is a crucial issue to get to an efficient and informative model and cannot be done by merely importing the concepts developed for physical experiments. This means providing efficient strategies for sampling the input space in order to get accurate predictions at untried inputs.

As suggested by the pioneers of the model-based CE, the output can be predicted by assuming joint Gaussian distribution of the responses with a covariance depending on the distance between the locations, as it was in the Kriging model, see Krige (1951) and Cressie (1986). The underlying principle is that the nearer an untried point is to the design points the better is the prediction. Based on this view, a good design strategy is to uniformly spread the points across the experimental region. That prompts to the use of the so called space-filling designs. Random designs, Stratified designs and Latin Hypercube designs are common choices, see e.g. McKay et al (1979), Fang et al (2006). The first two designs are not really satisfactory because they are not space filling marginally, i.e. in individual directions, and, moreover, are quite not satisfactory for global space filling especially for small number of design points. Whereas, even if the Latin Hypercube (LH) designs are not very satisfactory for space filling, they are satisfactory for space filling in individual directions, and this statement is valid for any number of design points. For a formal treatment of the subject see Welch et al (1992), Park (1994), Fang et al (2000), Butler (2001).

The content of the paper is as follows. We focus on the class of LH designs among the different space-filling ones. We want to investigate which LH designs have the best prediction features.

In Section 2, the Gaussian field is assumed to have distance-dependent covariance (stationarity) and it is defined on a subset of a regular lattice, i.e. the Cartesian

product of uniform one-dimensional lattices. This case correspond to specific applications, see e.g. Pistone and Vicario (2009), and it is specially fit for the methodology we use. The commonly used Euclidean distance is not natural when considering a regular lattice; therefore we switch to the Manhattan distance, as other authors have suggested, see Santner et al (2003, p. 138).

In Section 3, we assess the behavior of different LH designs by computing in closed form the Mean Square Prediction Error (MSPE) of the linear predictor. It should be noticed that the issue of identifiability which is of the highest importance in standard DoE is not relevant here.

Section 4 contains the main results of the paper. Specific study of LH design is performed through an example. We present a step by step discussion of LH designs with 2 factors with 4 levels each, including a full example of the use of the algebraic software.

In performing the aforementioned comparisons, there are two tricky computational problems: First, the computation of variance-covariance matrix of the design points and, Second, the computation of the closed form expression of the predictors variance. For the former problem, a solution is presented in a particular case of exponential correlation function, which is actually one of the most commonly used by the CE practitioners. For the latter, the difficulties are in the computation of the rational functions in the covariances with rational coefficients. In order to compute the predictor variances in closed form, we resort to the use of symbolic algebraic software such as CoCoA (Computations in Commutative Algebra), a freely available system for symbolic exact multivariate polynomial computation, see CoCoATeam. Other computations related with the exponential model for covariances are done with the software R, (<http://www.R-project.org/>). The final result is a general methodology to analyze the MSPE efficiency of classes of interest of training sets for regular lattices, with covariances of the form of an exponential of the Manhattan distance. All LHDs are classified according to the algebraic form of the predictor variance and the algebraic form of the determinant of the correlation matrix (entropy criteria). It is remarkable to observe that, at least in the examples we discuss, the two classifications coincide. This and other conclusions are discussed in Section 5.

2 Correlation function on a lattice

Let $Y(\mathbf{x})$ be a zero mean and stationary covariance Gaussian random field over a design space $\mathcal{X}_d \subset \mathbb{R}^d$, i.e.

$$\begin{aligned} E(Y(\mathbf{x})) &= 0, \\ \text{Cov}(Y(\mathbf{x}), Y(\mathbf{x} + \mathbf{h})) &= \sigma_Y^2 R(\mathbf{h}; \boldsymbol{\psi}), \end{aligned}$$

where σ_Y^2 is the field variance and R is its Stationary Correlation Function (SCF) depending only on the displacement vector \mathbf{h} between any pair of points in \mathcal{X}_d and on a vector parameter $\boldsymbol{\psi}$. A popular choice for the SCF is the power exponential

family

$$R(\mathbf{h}; \boldsymbol{\psi}) = \prod_{s=1}^d \exp(-\theta_s |h_s|^p) = \exp\left(-\sum_{s=1}^d \theta_s |h_s|^p\right), \quad 0 < p \leq 2, \quad (1)$$

where $\mathbf{h} = (h_1, \dots, h_d)$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$ are positive scale parameters, p is a common smoothing parameter, and $\boldsymbol{\psi} = (\boldsymbol{\theta}, p)$. The conditions on $\boldsymbol{\theta}$ are necessary and sufficient for (1) to be positive definite and, therefore, for the existence of a stationary Gaussian field with an SCF of that form, see Berg et al (1984). The original Krige's idea of a positive correlation between the outputs that decreases with increasing distance between their sites is true for the given SCF.

Our algebraic methodology is best described if we assume equal scale parameters $\theta_s = \theta$, $s = 1, \dots, d$. Moreover, it requires an integer value of p and the Gaussian field to be defined on a regular square lattice $\mathcal{X}_d = \{1, \dots, l\}^d$. As the Euclidean distance would not be really adapted to the lattice case, we restrict ourselves to the case $p = 1$, i.e. the Manhattan distance, $\|x - y\|_1 = \sum_{j=1}^d |x_j - y_j|$, as other authors have suggested, see Santner et al (2003, p. 138).

The SCF has the form

$$R(\mathbf{h}; \theta) = \prod_{s=1}^d \exp(-\theta |h_s|) = \exp\left(-\theta \sum_{s=1}^d |h_s|\right) = \exp(-\theta \|\mathbf{h}\|_1). \quad (2)$$

Let us first consider the univariate case $d = 1$. The distance function is $d_1(i, j) = |i - j|$, and the covariance function in algebraic form is

$$R_1(h; \theta) = \exp(-\theta d_1(i, j)) = t^{|i-j|}, \quad h = i - j, \quad i, j = 1, \dots, l,$$

where $t = \exp(-\theta) > 0$.

The matrix of the distances is

$$D_1 = \begin{matrix} & \begin{matrix} 1 & 2 & \dots & l \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ l \end{matrix} & \begin{bmatrix} 0 & 1 & \dots & l-1 \\ 1 & 0 & \dots & l-2 \\ \vdots & \vdots & \ddots & \vdots \\ l-1 & l-2 & \dots & 0 \end{bmatrix} \end{matrix}$$

and the covariance matrix is

$$\Gamma_1 = \begin{matrix} & \begin{matrix} 1 & 2 & \dots & l \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ l \end{matrix} & \begin{bmatrix} 1 & t & \dots & t^{l-1} \\ t & 1 & \dots & t^{l-2} \\ \vdots & \vdots & \ddots & \vdots \\ t^{l-1} & t^{l-2} & \dots & 1 \end{bmatrix} \end{matrix} \quad (3)$$

If $d = 2$, the distance function is (see Fig. 1):

$$d_2((i_1, i_2), (j_1, j_2)) = |j_1 - i_1| + |j_2 - i_2| = d_1(i_1, j_1) + d_1(i_2, j_2)$$

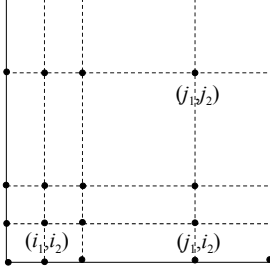


Fig. 1 A representation of the bivariate rectangular lattice $\mathcal{R}_2 = \{1, 2, \dots, l\}^2$

and the covariance function for $\mathbf{h} = \mathbf{i} - \mathbf{j} = (i_1 - j_1, i_2 - j_2)$ is:

$$R_2(\mathbf{h}; \theta) = \exp(-\theta d_2((i_1, i_2), (j_1, j_2))) = t^{|i_1 - j_1|} t^{|i_2 - j_2|}$$

It follows that the elements of the distance and covariance matrices, ordering rows and columns in lexicographic order, are given by:

$$D_2((i_1, i_2), (j_1, j_2)) = D_1(i_1, j_1) + D_1(i_2, j_2), \quad (4)$$

$$F_2((i_1, i_2), (j_1, j_2)) = F_1(i_1, j_1) \times F_1(i_2, j_2). \quad (5)$$

We write the matrix operations defined in (4) and (5) as $D_2 = D_1 \oplus D_1$ and $F_2 = F_1 \otimes F_1$, respectively. Note that D_2 and F_2 are $l^2 \times l^2$ matrices and \otimes denotes the Kronecker product of matrices, see e.g. Ortega (1987).

In generic dimension $d \geq 2$, we have the induction formulæ

$$D_d = D_{d-1} \oplus D_1, \quad F_d = F_{d-1} \otimes F_1. \quad (6)$$

The special algebraic structure of F_d allows us an algebraic treatment of the prediction problem. In particular, the computation of variances and covariances in closed form may be done by a symbolic algebraic software, so that we can compare the performance of each LH design on the basis of closed form algebraic expressions.

Indeed, the resulting statistical model is special. In fact, one can show by recursion that $\det F_1 = (1 - t^2)^{l-1}$ and that F_1^{-1} is tri-diagonal, e.g. for $l = 4$

$$F_1^{-1} = (1 - t^2)^{-1} \begin{bmatrix} 1 & -t & 0 & 0 \\ -t & 1 + t^2 & -t & 0 \\ 0 & -t & 1 + t^2 & -t \\ 0 & 0 & -t & 1 \end{bmatrix}$$

As F_1^{-1} is tri-diagonal, the 1-dimensional case is a Gaussian Markov Chain, precisely a discrete Ornstein-Uhlenbeck process, see (Santner et al 2003, p. 36). The d -dimensional case is a Kronecker product of Markov chains, therefore it is a Gaussian graphical model. We do not discuss further this interesting point and refer to Lauritzen (1996).

It has been observed by one of the anonymous referees that a Gaussian process with SCF (2) exhibits peculiar symmetries, precisely the joint distribution over the sites $\mathbf{x} \in \mathcal{X}_d$ depends only on the distances $\|\mathbf{x} - \mathbf{y}\|_1$, $\mathbf{x}, \mathbf{y} \in \mathcal{X}_d$, so that the distribution is invariant under any transformation of \mathcal{X}_d for which the Manhattan distance between points is an invariant. For example, the classification of subsets with 4 points of a 4×4 grid we shall obtain in Table 3 is based on the action of the dihedral group of the square. The method we present here does not require the preliminary study of such symmetries.

3 Predicting the output on lattice points

Sacks et al (1989a,b) suggested that the joint use of Kriging model as a meta-model together with Latin Hypercube designs as a training set is the option of choice in CE where no specific model is imposed by the application itself. The model considers the response $Y(\mathbf{x})$, for $\mathbf{x} \in \mathcal{X}_d \subset \mathbb{R}^d$, to be a realization of a Gaussian random field of the form

$$Y(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}),$$

where $\mathbf{f}'(\mathbf{x}) = [f_1(\mathbf{x}) \cdots f_m(\mathbf{x})]$ is a set of specified trend functions, $\boldsymbol{\beta}$ is a vector of parameters, and $Z(\mathbf{x})$ is a Gaussian random field with zero mean and SCF over \mathcal{X}_d . According to the Kriging's principle, the observed information on the random vector of the field variables $\mathbf{Y}^n = [Y(\mathbf{x}_1) \cdots Y(\mathbf{x}_n)]'$ at the training data set $\mathbf{x}_1, \dots, \mathbf{x}_n$ is used to predict the unobserved output at \mathbf{x}_0 . The underlying hypothesis, see Santner et al (2003), which is consistent with the covariance function in (1), assumes that the joint distribution of $Y(\mathbf{x}_0)$ and \mathbf{Y}^n is

$$\begin{bmatrix} Y(\mathbf{x}_0) \\ \mathbf{Y}^n \end{bmatrix} \sim N \left[\begin{bmatrix} \mathbf{f}'(\mathbf{x}_0) \\ \mathbf{F}^n \end{bmatrix} \boldsymbol{\beta}, \sigma_Z^2 \begin{bmatrix} 1 & \mathbf{r}_0' \\ \mathbf{r}_0 & \mathbf{R}^n \end{bmatrix} \right],$$

where \mathbf{F}^n is the $n \times m$ matrix with entries $f_j(\mathbf{x}_i)$, $i = 1, \dots, n$, $j = 1, \dots, m$, $\mathbf{r}_0' = [R(\mathbf{x}_0 - \mathbf{x}_1) \cdots R(\mathbf{x}_0 - \mathbf{x}_n)]$ is the correlation vector, and \mathbf{R}^n is the $n \times n$ correlation matrix whose (i, j) -element is $R(\mathbf{x}_i - \mathbf{x}_j)$, $i, j = 1, \dots, n$.

We consider in this paper the *ordinary* Kriging model, i.e. we assume $\mathbf{f}'(\mathbf{x})\boldsymbol{\beta} = \beta$, so that $Y(\mathbf{x}) = \beta + Z(\mathbf{x})$, where the trend is constant, and we assume the field variance to be unit, $\sigma_Z^2 = 1$, so that our model reduces to

$$\begin{bmatrix} Y(\mathbf{x}_0) \\ \mathbf{Y}^n \end{bmatrix} \sim N \left[\beta \mathbf{1}, \begin{bmatrix} 1 & \mathbf{r}_0' \\ \mathbf{r}_0 & \mathbf{R}^n \end{bmatrix} \right].$$

We use the Linear Unbiased Predictor (LUP). When the covariance is known, the Kriging methodology uses a linear spatial interpolation, i.e. the random variable $Y(\mathbf{x}_0)$ is predicted by an affine combination of the observed random variables,

$$\hat{Y}(\mathbf{x}_0) = a_0 + \sum_{i=1}^n a_i Y(\mathbf{x}_i). \quad (7)$$

The predictor $\hat{Y}(\mathbf{x}_0)$ is unbiased if and only if, for all β , $\beta = E(\hat{Y}(\mathbf{x}_0)) = a_0 + \sum_{i=1}^n a_i \beta$, i.e. $a_0 = 0$ and $\sum_{i=1}^n a_i = 1$. The predictor $\hat{Y}(\mathbf{x}_0) = \sum_{i=1}^n a_i Y_i$, $\sum_{i=1}^n a_i = 1$

is the Best LUP (BLUP) if the Mean Squared Prediction Error $\text{MSPE}[\hat{Y}(\mathbf{x}_0)] = E((\hat{Y}(\mathbf{x}_0) - Y(\mathbf{x}_0))^2)$ is minimized. If β is known, the BLUP is the conditional expectation of $Y(\mathbf{x}_0)$ given $\mathbf{Y}^n = [Y(\mathbf{x}_1) \cdots Y(\mathbf{x}_n)]'$,

$$\hat{Y}(\mathbf{x}_0) = \beta + \mathbf{r}'_0(\mathbf{R}^n)^{-1}(\mathbf{Y}^n - \beta \mathbf{1}) \quad (8)$$

and the value of the MSPE is $1 - \mathbf{r}'_0(\mathbf{R}^n)^{-1}\mathbf{r}_0$.

If β is to be estimated, the BLUP is given by (8) with β replaced by its generalized least squares estimator

$$\hat{\beta} = (\mathbf{1}'(\mathbf{R}^n)^{-1}\mathbf{1})^{-1}\mathbf{1}'(\mathbf{R}^n)^{-1}\mathbf{Y}^n$$

In such a case the MSPE, usually called Kriging variance, is larger because of an additional uncertainty component:

$$\begin{aligned} \text{MSPE}[\hat{Y}(\mathbf{x}_0)] &= (1 - \mathbf{r}'_0(\mathbf{R}^n)^{-1}\mathbf{r}_0) \\ &\quad + (1 - \mathbf{1}'(\mathbf{R}^n)^{-1}\mathbf{r}_0)'(\mathbf{1}'(\mathbf{R}^n)^{-1}\mathbf{1})^{-1}(1 - \mathbf{1}'(\mathbf{R}^n)^{-1}\mathbf{r}_0) \end{aligned}$$

In the next section we shall compute in closed form the MSPE as an algebraic function of the parameter $t = \exp(-\theta)$. Since the computational complexity of the symbolic algorithms is high, any reduction of the number of operation—however small—is of interest. To this aim, we derive now a more compact form of the BLUP (7) and of its MSPE, to be used in the following section.

Consider the linear transformation that maps the vector

$$[Y(\mathbf{x}_0) \ Y(\mathbf{x}_1) \ \cdots \ Y(\mathbf{x}_n)]'$$

to the vector $\tilde{\mathbf{Y}}$ whose entries are $\tilde{Y}_1 = Y(\mathbf{x}_0) - Y(\mathbf{x}_1)$, $\tilde{Y}_i = Y(\mathbf{x}_i) - Y(\mathbf{x}_1)$, $i \geq 2$. The generic LUP $\hat{Y}(\mathbf{x}_0) = \sum_{i=1}^n a_i Y(\mathbf{x}_i)$, $\sum_{i=1}^n a_i = 1$, can be written as

$$\begin{aligned} \hat{Y}(\mathbf{x}_0) &= a_1 Y(\mathbf{x}_1) + \sum_{i=2}^n a_i (\tilde{Y}_i + Y(\mathbf{x}_1)) \\ &= \left(\sum_{i=1}^n a_i\right) Y(\mathbf{x}_1) + \sum_{i=2}^n a_i \tilde{Y}_i \\ &= Y(\mathbf{x}_1) + \sum_{i=2}^n w_i \tilde{Y}_i, \end{aligned}$$

with unconstrained weights $w_i = a_i$, $i = 2, \dots, n$. If the w_i 's are such that $E(\tilde{Y}_1 | \tilde{Y}_i, i = 2, \dots, n) = \sum_{i=2}^n w_i \tilde{Y}_i$ the variance

$$\text{Var}\left(\tilde{Y}_1 - \sum_{i=2}^n w_i \tilde{Y}_i\right) = \text{Var}(Y(\mathbf{x}_0) - \hat{Y}(\mathbf{x}_0))$$

is minimized. Let

$$\tilde{\mathbf{R}} = \begin{bmatrix} \tilde{\mathbf{R}}_{11} & \tilde{\mathbf{R}}_{12} \\ \tilde{\mathbf{R}}_{21} & \tilde{\mathbf{R}}_{22} \end{bmatrix}$$

be the covariance matrix of $\tilde{\mathbf{Y}}$ partitioned $1|n-1$. We have $\mathbf{w} = \tilde{\mathbf{R}}_{12}\tilde{\mathbf{R}}_{22}^{-1}$ and

$$\text{MSPE}[\hat{Y}_0] = \tilde{\mathbf{R}}_{11} - \tilde{\mathbf{R}}_{12}\tilde{\mathbf{R}}_{22}^{-1}\tilde{\mathbf{R}}_{21}. \quad (9)$$

Table 1 The 24 LHDs for $l = 4$ and $d = 2$

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
11	11	11	14	14	11	11	11	13	13	13	14	14	13	13	13	12	12	12	14	14	12	12	12
22	22	24	21	21	24	23	23	21	21	24	23	23	24	22	22	23	23	24	22	22	24	21	21
33	34	32	32	33	33	34	32	32	34	31	31	32	32	34	31	31	34	33	33	31	31	34	33
44	43	43	43	42	42	42	44	44	42	42	42	41	41	41	44	44	41	41	41	43	43	43	44

Remark 1 The value of the parameter $t = \exp(-\theta)$ is usually estimated from the set of the training points, e.g. by Maximum Likelihood (or the restricted one), cross-validation or the posterior mode, and plugged in into the formula of the estimator. The estimated value is plugged into the BLUP formula: the final predictor is no more linear, even if it is still named Empirical Best Linear Unbiased Predictor (EBLUP). For a thoroughgoing reading, see Santner et al (2003, 64ff). In this paper, we do not consider this issue, because the LH designs are compared at each value of the t parameter in $[0, 1]$. Nevertheless, as the likelihood equations are algebraic, it is possible to use Algebraic Statistics to discuss such a case as it is done in Catanese et al (2006) for a different problem.

4 Latin Hypercube designs

A LH design (LHD) is a subset of the l^d lattice with l points that fully projects on each dimension. LHDs were introduced in CE by McKay et al (1979). There are $(l!)^{d-1}$ LHDs and their generation reduces to the generation of $d - 1$ permutations of $1, \dots, l$. Other types of fraction are much more difficult to sample from which partly explains the popularity of LHDs. See Table 1 for an example with $l = 4$ and $d = 2$. The point (i, j) is denoted by ij ; the LHD # 11 proceeds from the permutation $1234 \rightarrow 3412$. It has been observed that some of the LHDs are not attractive because they do not ensure a sufficient covering of the design space, see e.g. Ye et al (2000). The covering requirement depends on the closeness between the training points and prediction ones. The MSPE (9) is:

1. large when \mathbf{x}_0 is away from the training points $\mathbf{x}_1, \dots, \mathbf{x}_n$;
2. small when it is close to them;
3. zero at the experimental points because of the interpolatory property of Kriging.

Hence, a good planning should look for a specific choice of an LHD among the $(l!)^{d-1}$ that are available. The present paper is aimed to discuss a specific algebraic methodology to assess the properties of each LHD.

The computation of a BLUP from a LHD requires the inversion of a sub-matrix of Γ_d defined in (6). The sub-matrix has a special structure that we illustrate in the

case $l = 4, d = 2$. The full covariance matrix is

$$\Gamma_2 = \begin{matrix} & \begin{matrix} 11 & 12 & 13 & 14 & 21 & 22 & 23 & 24 & 31 & 32 & 33 & 34 & 41 & 42 & 43 & 44 \end{matrix} \\ \begin{matrix} 11 \\ 12 \\ 13 \\ 14 \\ 21 \\ 22 \\ 23 \\ 24 \\ 31 \\ 32 \\ 33 \\ 34 \\ 41 \\ 42 \\ 43 \\ 44 \end{matrix} & \begin{bmatrix} 1 & t & t^2 & t^3 & t & t^2 & t^3 & t^4 & t^2 & t^3 & t^4 & t^5 & t^3 & t^4 & t^5 & t^6 \\ t & 1 & t & t^2 & t^2 & t & t^2 & t^3 & t^3 & t^2 & t^3 & t^4 & t^4 & t^3 & t^4 & t^5 \\ t^2 & t & 1 & t & t^3 & t^2 & t & t^2 & t^4 & t^3 & t^2 & t^3 & t^5 & t^4 & t^3 & t^4 \\ t^3 & t^2 & t & 1 & t^4 & t^3 & t^2 & t & t^5 & t^4 & t^3 & t^2 & t^6 & t^5 & t^4 & t^3 \\ t & t^2 & t^3 & t^4 & 1 & t & t^2 & t^3 & t & t^2 & t^3 & t^4 & t^2 & t^3 & t^4 & t^5 \\ t^2 & t & t^2 & t^3 & t & 1 & t & t^2 & t^2 & t & t^2 & t^3 & t^3 & t^2 & t^3 & t^4 \\ t^3 & t^2 & t & t^2 & t^2 & t & 1 & t & t^3 & t^2 & t & t^2 & t^4 & t^3 & t^2 & t^3 \\ t^4 & t^3 & t^2 & t & t^3 & t^2 & t & 1 & t^4 & t^3 & t^2 & t & t^5 & t^4 & t^3 & t^2 \\ t^2 & t^3 & t^4 & t^5 & t & t^2 & t^3 & t^4 & 1 & t & t^2 & t^3 & t & t^2 & t^3 & t^4 \\ t^3 & t^2 & t^3 & t^4 & t^2 & t & t^2 & t^3 & t & 1 & t & t^2 & t^2 & t & t^2 & t^3 \\ t^4 & t^3 & t^2 & t^3 & t^3 & t^2 & t & t^2 & t^2 & t & 1 & t & t^3 & t^2 & t & t^2 \\ t^5 & t^4 & t^3 & t^2 & t^4 & t^3 & t^2 & t & t^3 & t^2 & t & 1 & t^4 & t^3 & t^2 & t \\ t^3 & t^4 & t^5 & t^6 & t^2 & t^3 & t^4 & t^5 & t & t^2 & t^3 & t^4 & 1 & t & t^2 & t^3 \\ t^4 & t^3 & t^4 & t^5 & t^3 & t^2 & t^3 & t^4 & t^2 & t & t^2 & t^3 & t & 1 & t & t^2 \\ t^5 & t^4 & t^3 & t^4 & t^4 & t^3 & t^2 & t^3 & t^3 & t^2 & t & t^2 & t^2 & t & 1 & t \\ t^6 & t^5 & t^4 & t^3 & t^5 & t^4 & t^3 & t^2 & t^4 & t^3 & t^2 & t & t^3 & t^2 & t & 1 \end{bmatrix} \end{matrix} \quad (10)$$

A generic LHD has the form $[1\sigma(1), \dots, n\sigma(n)]$ and the element of the sub-matrix with row name $i\sigma(i)$ and column name $j\sigma(j)$ is $t^{|j-i|+|\sigma(j)-\sigma(i)|} = t^{|j-i|} \times t^{|\sigma(j)-\sigma(i)|}$ therefore the sub-matrix is $\Gamma_1 \circ \sigma(\Gamma_1)$ where \circ denotes the component-wise (Hadamard) product of matrices and $\sigma(\Gamma_1)$ is the σ -permutation of rows and columns. For example, the correlation sub-matrix of the LHD # 11 is

$$\begin{matrix} & \begin{matrix} 13 & 24 & 31 & 42 \end{matrix} \\ \begin{matrix} 13 \\ 24 \\ 31 \\ 42 \end{matrix} & \begin{bmatrix} 1 & t^2 & t^4 & t^4 \\ t^2 & 1 & t^4 & t^4 \\ t^4 & t^4 & 1 & t^2 \\ t^4 & t^4 & t^2 & 1 \end{bmatrix} \end{matrix} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 1 & t & t^2 & t^3 \\ t & 1 & t & t^2 \\ t^2 & t & 1 & t \\ t^3 & t^2 & t & 1 \end{bmatrix} \end{matrix} \circ \begin{matrix} & \begin{matrix} 3 & 4 & 1 & 2 \end{matrix} \\ \begin{matrix} 3 \\ 4 \\ 1 \\ 2 \end{matrix} & \begin{bmatrix} 1 & t & t^2 & t \\ t & 1 & t^3 & t^2 \\ t^2 & t^3 & 1 & t \\ t & t^2 & t & 1 \end{bmatrix} \end{matrix}$$

4.1 The method

We present below a step-by-step description of our algebraic way to perform a comparison in the class of LHDs with d variables and/or factors, each one with l levels. Illustration is given for the case $d = 2, l = 4$. Computations are organized in a batch which calls the statistical software R (for the manipulation of the matrices) and the algebraic software CoCoA (for symbolic computations).

Global covariance matrix. The covariance matrix $\Gamma_d = \Gamma_1^{\otimes d}$, where Γ_1 is given in (3), is computed by Kronecker products, e.g. (10).

Generation of the LHDs. Generation of permutations of the integers $1, 2, \dots, l$ produces a $l \times (l!)^{d-1}$ table containing all the LHDs, e.g. Table 1.

Sub-setting. For each LHD and each \mathbf{x}_0 in its complement the joint covariance matrix is computed by sub-setting Γ_d ; MSPE for each \mathbf{x}_0 is symbolically computed as a rational function of t . The Total of the MSPEs over all points to be predicted (TMSPE) are obtained for all LHD as rational functions of t and subsequently TMSPEs are clustered according to the algebraic form of the rational function. The TMSPE index replaces the Integrated MSPE in the discrete case. For example, with $l = 4$ and $d = 2$, the following output is obtained from CoCoA. It is a list of rational functions, each one representing the MSPE of one of the seven classes.

Table 2 The 24 LHDs for $l = 4$ and $d = 2$ classified according the TMSPE criterion (bottom line)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
11	11	11	14	14	11	11	11	13	13	13	14	14	13	13	13	12	12	12	14	14	12	12	12
22	22	24	21	21	24	23	23	21	21	24	23	23	24	22	22	23	23	24	22	22	24	21	21
33	34	32	32	33	33	34	32	32	34	31	31	32	32	34	31	31	34	33	33	31	31	34	33
44	43	43	43	42	42	42	44	44	42	42	41	41	41	44	44	41	41	41	43	43	43	43	44
1	2	3	4	3	4	3	5	3	6	7	2	1	2	3	4	3	4	3	5	3	6	7	2

%% C1:

$$(2t^8 - 4t^7 - 12t^6 + 4t^5 + 8t^4 + 20t^3 + 12t - 30)/(t^4 - t^2 - 2)$$

C2:

$$(t^{13} + t^{12} + 1/2t^{11} - 11/2t^{10} - 31/2t^9 - 19/2t^8 - 5/2t^7 + 71/2t^6 + 43/2t^5 + 71/2t^4 + 13t^3 - 27t^2 - 18t - 30)/(t^7 + t^6 - 2t^4 - 2t^3 - 4t^2 - 2t - 2)$$

C3:

$$(1/2t^{15} + 1/2t^{14} + t^{13} - 4t^{12} + 3/2t^{11} - 17/2t^{10} - 47/2t^9 + 9/2t^8 - 7t^7 + 48t^6 + 59/2t^5 + 49/2t^4 + 16t^3 - 35t^2 - 18t - 30)/(t^8 + t^7 - 4t^4 - 2t^3 - 5t^2 - 2t - 2)$$

C4:

$$(1/2t^{14} - 1/2t^{12} - 2t^{11} + 3t^{10} - 4t^9 - 18t^8 + 12t^7 - 27/2t^6 + 40t^5 + 3/2t^4 + 38t^3 - 39t^2 + 12t - 30)/(t^6 - t^4 - 4t^2 - 2)$$

C5:

$$(t^{15} + t^{14} + t^{13} - 7t^{12} - 5t^{11} - 5t^{10} - 23t^9 + 13t^8 + 2t^7 + 42t^6 + 20t^5 + 20t^4 + 22t^3 - 34t^2 - 18t - 30)/(t^9 + t^8 + 2t^7 - 2t^6 - 4t^4 - t^3 - 5t^2 - 2t - 2)$$

C6:

$$(t^9 - 3t^8 + 2t^7 + 2t^6 - 22t^5 + 8t^4 + 28t^3 + 8t^2 - 9t - 15)/(t^3 - t^2 - t - 1)$$

C7:

$$(3t^8 - 2t^7 + 6t^6 - 9t^5 - 5/2t^4 - 10t^3 + 10t^2 - 3t + 15/2)/(t^2 + 1/2)$$

which, in turn, gives rise to the classification shown in the last row of Table 2.

Performance evaluation Each rational function is studied to assess the performance in each class of LHD. For example, in graphical form as in Fig. 2; the classification obtained is best appreciated by looking at a graphical representation of the 24 LHDs in Table 3. Each of the 7 classes is an orbit of the action of the dihedral group of the square on the set of all LHDs, because of the dependence of the joint distribution of the Gaussian field on L^1 pairwise distance between points.

4.2 Discussion of the example

In this case there is not much difference between that the LHDs if the criteria is TMSPE. Some difference can be seen in the second graph showing relative difference near $t = 1$, that is $\theta = 0$ (constant covariance). As the computations are symbolic there is no risk of numerical errors near the critical point. Other criteria have been suggested, for example based on entropy, see Shewry and Wynn (1987). Our methodology, when applied to the entropy instead of TMSPE, produces the same classifica-

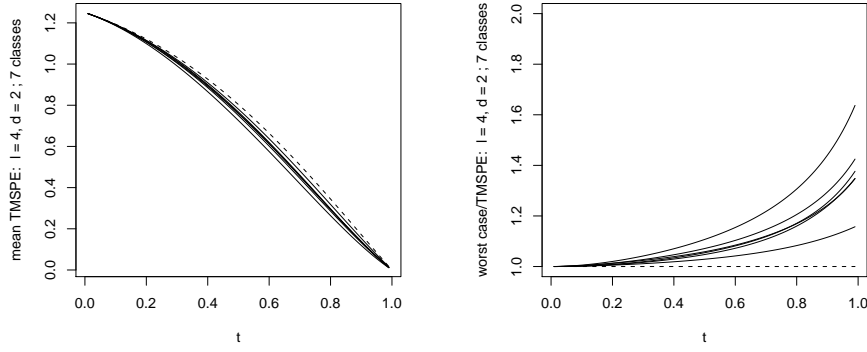


Fig. 2 Performance of LHDs for $l = 4$ and $d = 2$

Table 3 24 LHDs classified according the TMSPE criteria

Class	
1	
2	
3	
4	
5	
6	
7	

tion of LHDs as in Table 2, because of the invariance argument, and the same ranking of classes.

Class 6 shows the best performance. It consists of LHDs which are maximin in the L^1 distance, see Morris and Mitchell (1995); it also consists of “tilted” 2^2 designs; the designs in this class are called U-design in Tang (1993). Classes 3, 4, 5, are essentially equivalent and worse than class 6. Designs in class 5 are called cyclic designs in Bates et al (1996). Class 2 is second worse. Classes 1 and 4 consist of regular fractions 4^{2-1} . Class 3 is made of regular fractions 2^{4-2} (pseudo factors).

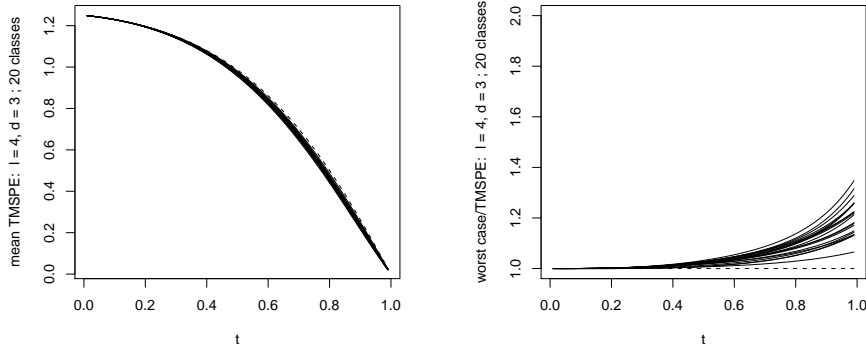


Fig. 3 Performance of LHDs for $l = 4$ and $d = 3$

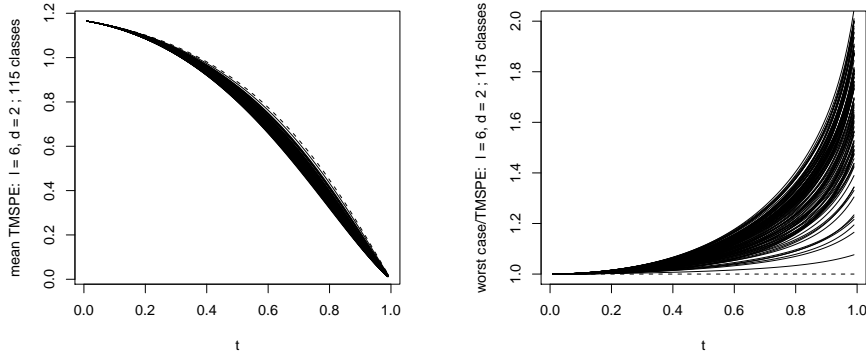


Fig. 4 Performance of LHDs for $l = 6$ and $d = 2$

4.3 Other examples

The methodology we have illustrated in the case $l = 4$ and $d = 2$ is, in principle, of general applicability. Fig. 3 shows the results in the case $l = 4$ and $d = 3$, and Fig. 4 the results for $l = 6$ and $d = 2$. The choice of these dimensions for levels and number of factors is merely indicative of the type of comparisons that are available. The dashed lines represent the TMSPE of the LHDs whose points lie on the diagonals of the grid. These designs are considered not properly space-filling, even if they have nice marginal properties and the largest value of the respective TMSPE confirm their poor capability of prediction. Other remarkable features shown are the minimum value of the TMSPE corresponding to maximum correlation $t = 1$ and the maximum value of the TMSPE corresponding to the null correlation $t = 0$ between the training points (i.e. independence, according to the assumption of normality).

This is coherent with the Kriging prediction methodology: predicting at an untried location, observations closer to it should influence more the prediction because of the existing correlation. In the right-hand graphs, we resort to a different representation of the comparison index relating the TMSPEs to the worst case, in order to magnify the differences: the ratios between the TMSPE of the diagonal LH designs and the ones corresponding to each of the other classes are plotted vs $t = \exp(-\theta)$.

5 Discussion

The main result of this paper is the illustration of symbolic computations in a particular model of Kriging. It shows the potential applicability, in this area, of ideas from Algebraic Statistics, see. the review in Gibilisco et al (2009). Interesting exploratory results are obtained, for example the neat classification and ranking of the 4×4 LHDs shown in Table 2, Fig. 2, and Table 3. The same classification and ranking was obtained for other optimality criteria, e.g. entropy. While the classes depend on the symmetry of the problem under the dihedral group of the square, it is notable that it is the result of a symbolic computation. The algebraic form is also of use in evaluating the relative error near the critical point $t = 1$.

The methodology depends on the algebraic form of the SCF and could be implemented for any class of training set, LHDs being an example. Some of the other popular performance criteria are based on non-algebraic operations, typically max MSPE at untried points, and cannot be solved algebraically in a standard way.

This methodology has limitations when the number of levels and/or factors is large. Indeed, the very high complexity of the symbolic computations limits considerably the dimension of the problems that can be practically solved. In this paper, we have not tried any optimization of the program we have used, which consists of a sequence on runs from standard packages, nor we considered writing any ad-hoc software. Such an optimization would require the study of efficient ways to generate the training sets from the class of interest, which is a general problem not depending on the algebraic methodology, and the study of algorithms for the symbolic computation of the inverse and the determinant of a correlation matrix with polynomial entries.

Acknowledgments A preliminary versions of the present paper was presented at the EMBIS-EMSE 2009 Conference, July 1-3, Saint-Étienne, France. We thank the editors and anonymous referees for comments and suggestions.

References

- Bates RA, Buck RJ, Riccomagno E, Wynn HP (1996) Experimental designs and observations for large systems. *Journal of the Royal Statistical Society B* 58:77–94
- Berg C, Christensen JPR, Ressel P (1984) Harmonic analysis on semigroups. Theory of positive definite and related functions, Graduate Texts in Mathematics, vol 100. Springer-Verlag, New York
- Butler NA (2001) Optimal and orthogonal Latin Hypercube designs for computer experiments. *Biometrika* 88(3):847–857
- Catanese F, Hoşten S, Khetan A, Sturmfels B (2006) The maximum likelihood degree. *Amer J Math* 128(3):671–697

- CoCoATeam (online) CoCoA: a system for doing Computations in Commutative Algebra. Available at <http://cocoa.dima.unige.it>
- Cressie N (1986) Kriging nonstationary data. *Journal of the American Statistical Association* 81(395):625–634
- Fang KT, Lin DKJ, Winker P, Zhang Y (2000) Uniform design: theory and application. *Technometrics* 42(3):237–248
- Fang KT, Li R, Sudjianto A (2006) Design and modeling for computer experiments. Computer Science and Data Analysis Series, Chapman & Hall/CRC, Boca Raton, FL
- Gibilisco P, Riccomagno E, Rogantin M, Wynn HP (eds) (2009) Algebraic and Geometric Methods in Statistics. Cambridge University Press
- Krige DG (1951) A statistical approach to some basic mine valuation problems on the witwatersrand. *J of the Chem, Metal and Mining Soc of South Africa* 52(6):119139
- Lauritzen SL (1996) Graphical models. The Clarendon Press Oxford University Press, New York, Oxford Science Publications
- McKay MD, Beckman RJ, Conover WJ (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21:239–245
- Morris MD, Mitchell TJ (1995) Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference* 43(3):381–402
- Ortega JM (1987) Matrix theory. The University Series in Mathematics, Plenum Press, New York, a second course
- Park JS (1994) Optimal Latin-hypercube designs for computer experiments. *J Statist Plann Inference* 39(1):95–111
- Pistone G, Vicario G (2009) Statistics for Innovation. Statistical Design of “Continuous” Product Innovation, Springer-Verlag, chap Random Generation of Designs in Computer Experiments, pp 91–102
- Qian PZG, Wu H, Wu CFJ (2008) Gaussian process models for computer experiments with qualitative and quantitative factors. *Technometrics* 50(3):383–396, DOI 10.1198/004017008000000262, URL <http://dx.doi.org/10.1198/004017008000000262>
- Sacks J, Schiller SB, Welch WJ (1989a) Designs for computer experiments. *Technometrics* 31(1):41–47
- Sacks J, Welch WJ, Mitchell TJ, Wynn HP (1989b) Design and analysis of computer experiments (with discussion). *Statistical Science* 4:409–435
- Santner TJ, Williams BJ, Notz WI (2003) The design and analysis of computer experiments. Springer Series in Statistics, Springer-Verlag, New York
- Sasena MJ (2002) Flexibility and efficiency enhancements for constrained global design optimization with kriging approximations. PhD thesis, University of Michigan
- Shewry MC, Wynn HP (1987) Maximum entropy sampling. *Journal of Applied Statistics* 14:165–170
- Tang B (1993) Orthogonal array-based Latin hypercubes. *J Amer Statist Assoc* 88(424):1392–1397
- Welch WJ, Buck RJ, Sacks J, Wynn HP, Mitchell TJ, Morris MD (1992) Screening, predicting, and computer experiment. *Technometrics* 34(1):15–25
- Ye KQ, Li W, Sudjianto A (2000) Algorithmic construction of optimal symmetric Latin hypercube designs. *J Statist Plann Inference* 90(1):149–159